

ОБЩИЕ ПРИНЦИПЫ ОРГАНИЗАЦИИ ВНУТРИВУЗОВСКОГО КОНТРОЛЯ И УПРАВЛЕНИЯ КАЧЕСТВОМ ОБРАЗОВАТЕЛЬНЫХ ТЕСТОВЫХ МАТЕРИАЛОВ НА ЭТАПЕ ИХ РАЗРАБОТКИ

И.Ю. Петрова, С.В. Окладникова

В статье рассматриваются особенности создания образовательных тестовых материалов с применением автоматических и автоматизированных методов, необходимость оценки качества тестовых материалов на этапе их разработки. Определены основные ошибки, оказывающие влияние на качество разрабатываемых тестовых материалов. Предложены принципы организации контроля и управления качеством тестовых материалов внутри вуза

Ключевые слова: *тестовый материал, сложность формулировки, критерий качества*
Keywords: *test material, complexity language, performance criterion*

В настоящее время в вузах компьютерное тестирование (КТ) как метод оценки уровня знаний учащихся приобретает все большую популярность. Однако объективность выставления оценки напрямую зависит от качества используемого тестового материала (ТМ).

Содержание ТЗ, применяемых в КТ, может быть представлено в формате текста, статической графики, мультимедиа (т.е. динамической графики и/или аудио), а также их различными сочетаниями друг с другом. Анализ ТМ, накопленных в Астраханском государственном университете (АГУ) в период за 2004–2008 гг., показал, что ТЗ с формулировками вопроса и ответов в текстовом формате преобладают и составляют 84 % (рис. 1) от общего количества.

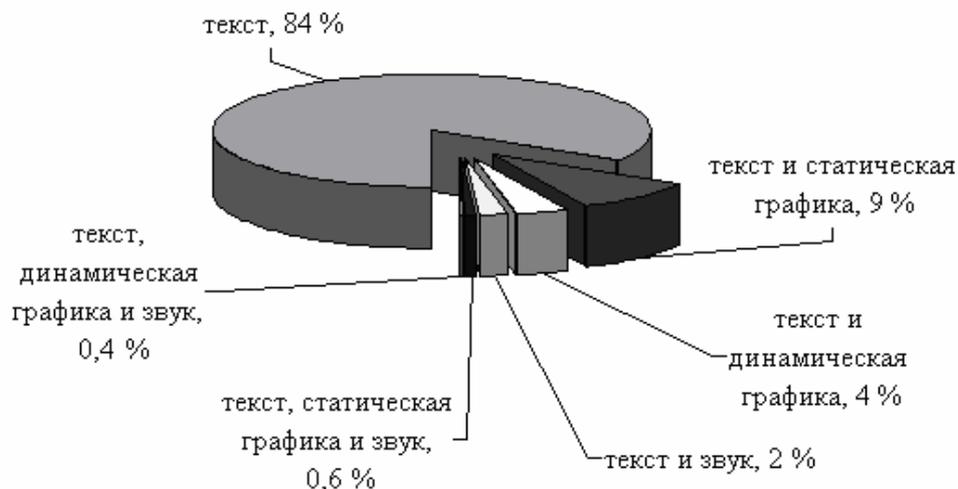


Рис. 1. Соотношение видов наполнения ТЗ в общем массиве тестовых материалов, накопленных в АГУ за 2004–2007 гг.

Преобладание ТЗ с текстовым содержанием формулировок или с дополнительными элементами статической графики объясняется простотой в их технологическом исполнении. Сегодня разработка ТЗ рассматриваемых форматов, как правило, осуществляется с помощью либо встроенных в конструкторы тестов редакторов, либо с помощью других текстовых редакторов (например, MS Word и т.п.) и не вызывает особых затруднений. Разработка же объектов динамической графики и аудио- объектов, используемых в содержании ТЗ, представляется нетривиальной задачей, требующей от разработчиков не только дополнительных знаний по основам работы в графических, мультимедиа, видео- и аудио- редакторах (Photoshop, Flash, 3D Max и т.п.) но и дополнительных материальных затрат, связанных с их приобретением.

Традиционно для разработки ТМ используются *автоматизированные* методы, в которых содержание и форма ТЗ определяются автором «вручную», а затем фиксируются на электронном носителе с помощью редакторов (например, Блокнот, MS Word, Paint, Photoshop и т.п.) или существующего в АСТ конструктора тестов. В процессе формирования содержания ТЗ авторами могут быть допущены ошибки, снижающие качество разрабатываемых ТМ. Например, ошибки:

1) технические, связанные с правилами оформления ТЗ, которые существуют в эксплуатируемой АСТ (неверно указан тип вопроса, отсутствует указатель правильности ответа, дублируются формулировки и ответы ТЗ и т.п.);

2) грамматические, искажающие смысл задания или делающие невозможным его понимание;

3) методические, при которых нарушаются основные методические требования, предъявляемые к содержанию ТЗ (корректность содержания формулировок вопросов и ответов ТЗ, легкость чтения; исключение в вариантах ответов ТЗ частицы «не», отрицающих предлогов и слов и т.п.);

4) технологические, связанные со спецификой выбранной формы ТЗ (общее количество вариантов ответов и дистракторов, пар соответствий и т.п.);

5) содержательные, связанные с нарушением принципов отбора содержания ТЗ.

Анализ существующих образовательных ТМ, разработанных преподавателями на базе российских вузов, показал, что наиболее распространенными ошибками, допускаемыми авторами, являются методические и технологические (рис. 2).

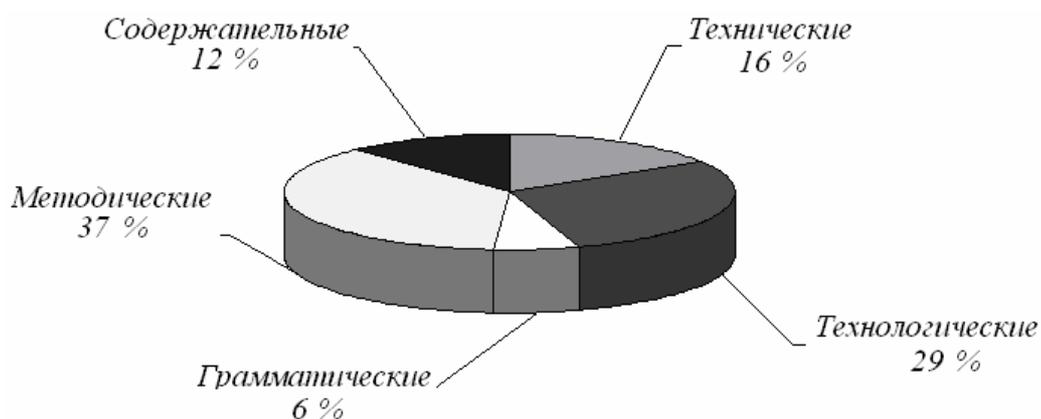


Рис. 2. Соотношение допускаемых преподавателями ошибок в отношении разрабатываемых ими тестовых заданий, %

Автоматические методы ориентированы на генерацию содержания ТЗ (преимущественно текстового типа) при наличии некоторой электронной базы знаний (БЗ) соответствующей предметной области (ПО). Активное использование этих методов позволило бы сократить ряд ошибок, однако недостаточно развитый на сегодняшний день математический аппарат компьютерной лингвистики использует формальный подход к формированию содержания ТЗ. Вследствие этого тестовые вопросы, составленные *автоматическими* методами, позволяют проверить те знания испытуемых (например, знание понятий, определений, терминов, фактов, формул и т.п.), которые были отображены в используемой БЗ.

Оценка качества ТМ в настоящее время выполняется совокупностью *экспериментально-статистических* и *педагогических* методов.

Особенностью *экспериментально-статистических* методов является необходимость в накоплении больших объемов статистической информации результатов тестирования, что для ряда дисциплин, преподаваемых в российских вузах, является проблематичным в силу лимита учебных часов, отведенных (в соответствии с учебным планом специальности) на их изучение, или в силу малочисленности группы учащихся. В этом случае накопление необходимого количества статистики будет происходить в течение нескольких лет, что, в свою очередь, может привести к потере актуальности оцениваемого ТМ. При этом обработка полученных результатов тестирования осуществляется с использованием сложного математи-

ко-статистического аппарата и требует наличия специальных программных продуктов, так как обработка «вручную» или с частичным привлечением технических средств потребует от автора ТМ, во-первых, изучения используемых методов оценки качества ТМ (чтобы правильно интерпретировать полученные результаты анализа), а во-вторых, проведения большой расчетной и вычислительной работы.

Использование *экспериментально-статистических* методов ориентировано в основном на анализ качества теста. Под «тестом» понимается система ТЗ, обеспечивающая качество и эффективность педагогического измерения. При этом процесс отбора «некачественных» ТЗ не позволяет автору определить возможную причину «выбраковки» такого ТЗ и указать тип ошибки с целью ее дальнейшего исправления. Поэтому для контроля качества содержания ТЗ в ряде вузов используются *педагогические методы*, основанные на экспертной оценке.

При оценке качества ТМ (разработанных на базе учебных заведений) с использованием *педагогических* методов существенной оказывается проблема, связанная с определением численности экспертной группы. Как правило, в качестве экспертов выступают преподаватели того же вуза, являющиеся либо специалистами соответствующей ПО, либо родственной с ней. Однако часто ТМ разрабатываются по конкретным (узконаправленным) разделам учебной дисциплины. Вследствие существующей специфики отдельных дисциплин и дефицита в вузах специалистов по этим дисциплинам возникает проблема с подбором экспертов. В результате состав экспертной группы оказывается малочисленным и не позволяет обеспечивать достаточную статистическую достоверность выборочной оценки, так как при небольшом числе экспертов на общую групповую оценку существенное влияние оказывают индивидуальные предпочтения экспертов. С целью увеличения численности экспертной группы было бы рациональным привлечение внешних специалистов (например, преподавателей других вузов). Однако при этом возникают дополнительные трудности организации экспертного опроса, связанные с увеличением затрат времени и денежных средств на проведение экспертизы.

Организация внутривузовского процесса управления качеством ТМ предполагает проведение следующих взаимодополняющих видов контроля (рис. 3):

1) *первичный* – контроль, выполняемый автором в отношении всех видов ошибок на этапе формирования содержания ТЗ до ввода их в АСТ. При этом проверка грамматических ошибок может быть выполнена автоматизировано в случае, если при разработке содержания ТЗ использовались текстовые редакторы типа MS Word, обеспечивающие поддержку правописания. Остальные ошибки, как правило, проверяются автором «визуально»;

2) *технический* – контроль, выполняемый автором в процессе ввода ТЗ в базу АСТ в соответствии с принятой в ней формой записи данных, а также реализованной компьютерной поддержкой разработчика ТМ. ТЗ, содержащие технологические ошибки, однозначно не вводятся в АСТ и подлежат обязательной корректировке (которая выполняется автором) с последующим повторным вводом. ТЗ, содержащие остальные виды ошибок, вводятся в АСТ, при этом система автоматически исправляет их или формирует соответствующие рекомендации. Обнаружение и исправление ошибок зависит от реализованных в АСТ алгоритмов контроля и управления различными видами ошибок;

3) *внутривузовская экспертиза* проводится группой экспертов (или одним экспертом) на завершающем этапе разработки ТЗ в отношении методических и содержательных ошибок, допущенных автором, и предполагает утверждение ТМ на учебно-методическом совете вуза (или на заседании кафедры) с выдачей соответствующих рецензий. Для ТМ, не прошедших экспертизу, заключение эксперта содержит соответствующие рекомендации по их доработке.

правописания, которая основана на проведении *лексического, морфологического и семантического* анализов. Реализация данного вида контроля осуществляется существующими методами компьютерной лингвистики.

Лексический анализ предполагает проверку корректности ввода отдельных символов путем их сравнения с символами, заранее определенными в массиве допустимых символов (МДС). В качестве допустимых символов, используемых при наборе текстовых формулировок ТЗ, можно рассматривать все буквы русского и английского алфавитов, цифры и другие служебные символы, например, №, *, % и т.п.

Основой для проверки орфографии и пунктуации выступают различные методы морфологического и синтаксического анализа русского языка.

Морфологический анализ связан с анализом отдельных слов и основан на использовании морфологических словарей или проведении морфемного анализа.

Синтаксический анализ связан с анализом отдельных предложений и позволяет определять взаимосвязи отдельных слов и частей предложений. Отдельные слова представляют собой вершины графа, которые связаны между собой определенным образом. На основе системы связей разрабатываются представления синтаксической структуры предложения. При этом формируется четкая система правил, согласно которым производится проверка пунктуации.

Для обеспечения автора ТМ автоматизированной поддержкой в отношении допускаемых им при создании ТЗ методических и технологических ошибок автором был разработан алгоритм управления фактической сложностью формулировок ТЗ.

Алгоритм управления сложностью формулировок ТЗ

1. Расчет фактических числовых значений частных критериев (C_1-C_6) оценки сложности формулировок создаваемого ТЗ.

2. Расчет фактического значения интегрального критерия (I_n) оценки сложности формулировки создаваемого ТЗ.

3. Определение максимально допустимого значения интегрального критерия (I_{max}) в зависимости от указанной автором ПО дисциплины, по которой данное ТЗ разрабатывается.

4. Сравнение фактического значения интегрального критерия (I_n) с предельно-допустимыми (I_{min}) и (I_{max}) .

5. Если $(I_{min}) < (I_n) < (I_{max})$, то сложность формулировки введенного ТЗ считается допустимой, содержание ТЗ сохраняется в базе АСТ и выполняется переход к шагу 9; в противном случае переходим к шагу 6.

6. Сравнение фактических значений частных критериев (C_1-C_6) с их предельно-допустимыми значениями $(C_i)_{mix}$ и $(C_i)_{max}$.

7. Если $(C_i)_{mix} < (C_i) < (C_i)_{max}$, то значение критерия считается допустимым; в противном случае автору выдаются рекомендации о необходимости изменения рассматриваемого критерия (C_i) .

8. Режим редактирования формулировки данного ТЗ.

9. Режим ввода нового ТЗ.